



Working Paper 04-11
Statistics and Econometrics Series 04
February 2004

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

A RANGE UNIT ROOT TEST

Felipe M. Aparicio^{*}
Alvaro Escribano
Ana García

Abstract

Since the seminal paper by Dickey and Fuller in 1979, unit-root tests have conditioned the standard approaches to analyse time series with strong serial dependence, the focus being placed in the detection of eventual unit roots in an autorregressive model fitted to the series. In this paper we propose a completely different method to test for the type of “long-wave” patterns observed not only in unit root time series but also in series following more complex data generating mechanisms. To this end, our testing device analyses the trend exhibit by the data, without imposing any constraint on the generating mechanism. We call our device the *Range Unit Root (RUR) Test* since it is constructed from *running ranges* of the series. These statistics allow a more general characterization of a strong serial dependence in the mean behavior, thus endowing our test with a number of desirable properties, among which its error-model-free asymptotic distribution, the invariance to nonlinear monotonic transformations of the series and the robustness to the presence of level shifts and additive outliers. In addition, the RUR test outperforms the power of standard unit root tests on near-unit-root stationary time series and is asymptotically immune to noise.

Keywords: Unit roots tests; Structural breaks; Nonlinearity; Additive outliers, Near-unit root time series; Invariance; Robustness; Running ranges, Noise immunity.

^{*}Aparicio, Statistics and Econometrics Department, University Carlos III of Madrid, Spain; Escribano, Economics Department, University Carlos III of Madrid, Spain; García, Statistics and Econometrics Department, University Carlos III of Madrid, Spain. The authors are grateful to Daniel Peña, Juan Romo, Esther Ruiz and Ismael Sánchez from Universidad Carlos III de Madrid, to Dag Tjøstheim and Jostein Paulsen from the University of Bergen, to Jorg Breitung from Humboldt University in Berlin, and to Javier Villarreal from University of Salamanca, for their useful comments. We acknowledge financial support from the Spanish CYCIT under grant PB98-0030.

A RANGE UNIT ROOT TEST

Felipe M. Aparicio
(Corresponding author: aparicio@est-econ.uc3m.es)
Dept. of Statistics and Econometrics
Universidad Carlos III de Madrid

Alvaro Escribano
Dept. of Economics
Universidad Carlos III de Madrid and Georgetown University

Ana García
Dept. of Statistics and Econometrics
Universidad Carlos III de Madrid

November, 2003

Abstract

Since the seminal paper by Dickey and Fuller in 1979, unit-root tests have conditioned the standard approaches to analyse time series with strong serial dependence in the mean behavior, the focus being placed in the detection of eventual unit roots in an autorregressive model fitted to the series. In this paper we propose a completely different method to test for the type of long-wave patterns observed not only in unit root time series but also in series following more complex data generating mechanisms. To this end, our testing device analyses the trend exhibited by the data, by imposing very few constraints on the generating mechanism. We call our device the Range Unit Root (RUR) Test since it is constructed from the running ranges of the series. These statistics endow the test with a number of desirable properties, among which its error-model-free asymptotic null distribution, the invariance to monotonic transformations of the series, and the robustness to the presence of parameter shifts and additive outliers. Moreover, the RUR test outperforms the power of standard unit root tests on near-unit-root stationary time series and is asymptotically immune to noise.

Key Words: Unit Roots Tests, Structural Breaks, Nonlinearity, Additive Outliers, Near-Unit Root Time Series, Invariance, Robustness, Running Ranges, Noise Immunity.

1 Introduction

Many overwhelming low-frequency non-periodic components in time series are associated with the presence of unit roots in their Data Generating Process (DGP). Such time series are said to be integrated. The pioneering work of Nelson and Plosser (1982) led to the belief that many economic time series were best described in this way. This prompted a large amount of research on unit root time series, covering both theoretical and empirical aspects. The unit root paradigm has important practical implications since it entails that shocks have a permanent effect on a variable, or equivalently that the fluctuations they cause are not transitory.

The existence of unit roots in time series is investigated by means of unit root tests. The application of standard unit root tests, such as the Dickey-Fuller (DF hereafter) test (Dickey and Fuller, 1979), has been an important step in the construction of a useful parametric model for many economic time series. In a one-sided DF test, the null hypothesis of a unit root in a series x_t , say $H_0 : (1 - B)(x_t - \bar{x}_t) = \varepsilon_t$, is tested against the alternative $H_1 : (1 - \beta B)(x_t - \bar{x}_t) = \varepsilon_t$ with $|\beta| < 1$, where \bar{x}_t denotes the mean of x_t . If the alternative is rejected then x_t is supposed to follow a unit root time series model.

Unit root time series models impose, however, severe restrictions on the DGP's of the data. Many real world time series exhibit nonlinearities, outliers, and structural breaks either in the mean or in the variance. All these features, which cannot be properly captured with random-walk-like models, fool standard unit root tests (see for instance, Granger and Hallman, 1991, and Ermini and Granger, 1993).

Alternative procedures for testing unit roots were proposed by Lo (1991), Kwiatkowski et al. (1992), Stock (1994) and Bai and Perron (1998). Yet, they were reported to have poor power performances when confronted to deviations from the standard linear context (see for example, Sims, 1988; Perron, 1989; Perron, 1990; Schotman and Van Dijk, 1991; 1995a).

The appropriate handling of such departures as parameter shifts, trend breaks and nonlinearities calls for the development of robust unit root tests. The rejection of the unit root hypothesis by standard tests together with the acceptance of a wider null by a robust testing procedure may be more than enough to justify the search for a competing nonlinear time series model.

In general, the nonparametric avenue of research have deserved little attention in unit root testing problem. As an example, Burrridge and Guerre (1996) proposed a nonparametric unit root testing device based on the different behavior of the level crossings of stationary series and random walks. However the rather poor results reported by these authors and their subsequent gloomy conclusions may have discouraged further research on such approaches.

In this paper, we try to provide evidence of the opposite by presenting a nonparametric Range Unit Root (RUR hereafter) test whose superiority with respect to the standard approaches is remarkable. First, it is invariant to monotonic transformations and to the distribution of the model errors. Second it is robust against structural breaks, parameter shifts and additive outliers. Third,

it does not depend on the variance of any stationary alternative and thereby outperforms standard tests also in terms of power on near-unit-root stationary time series. Finally, the RUR test is not affected asymptotically by the presence of additive noise on the series.

The RUR test is a natural follow-up of the methodology proposed in Aparicio (1995a), in Aparicio and Granger (1995b) and in Aparicio, Escribano and Garcia (2000) for robustizing cointegration tests, recently extended to unit root testing with some empirical applications in Aparicio, Escribano and Garcia (2002). The structure of the paper is as follows. In Section 2 we explain the heuristics which motivate the proposed methodology. This will lead us to define the RUR test in Section 3, where we also discuss its small sample behavior under the null hypothesis of a single unit root. Section 5 deals with the asymptotic null distribution of the test while Section 6 studies its power performances and its consistency against both stationary, integrated and trending alternatives. Section 7 analysis the behavior of test under different departures from the standard unit root tests' assumptions. Finally, after the concluding remarks in Section 8, an appendix is devoted to the proofs of the main theoretical results.

2 Time Series Analysis based on Ranges

Many time series not generated by unit-root models exhibit similar mean behavior to those which are. The objective of this section is to investigate alternative procedures for assessing the presence of unit-root like features, not necessarily caused by unit roots. We will begin by studying the behavior of the sequence of running ranges in both stationary and random walk time series.

The range of a data sample is defined in terms of its extremes. Formally, for a given time series x_t , the statistics $x_{1:i} = \min\{x_1, \dots, x_i\}$ and $x_{i:n} = \max\{x_1, \dots, x_i\}$ are called the i -th extremes (see for instance Galambos, 1984). When the sample comes from a time series x_t , a monotonically increasing sequence of ranges can be obtained as $R_i^{(x)} = x_{i:n} - x_{1:i}$, for $i = 1, 2, 3, \dots, n$, where n denotes the sample size. The total number of "new extremes" or records in a sample of size n is given by the quantity $\sum_{i=1}^n 1(\Phi R_i^{(x)} > 0)$; where $1(\cdot)$ is the indicator function.

It can be shown that the long-run frequency of new records, $n^{-1} \sum_{i=1}^n 1(\Phi R_i^{(x)} > 0)$; vanishes faster for stationary time series than for series containing a unit root -these latter series are often referred to as integrated "of order 1", or briefly as $I(1)$: In particular, for i.i.d. sequences of random variables we have (see for instance Embrechts, Klüppelberg and Mikosch, 1999):

$$n^{-1} \sum_{i=1}^n 1(\Phi R_i^{(x)} > 0) = O(n^{-1} \log n):$$

This result still holds for stationary series satisfying the so-called "Berman condition", which requires the covariance sequence of the series $\gamma_i = \text{Cov}(x_t, x_{t+i})$ to satisfy

to decrease faster than $(\log i)^{i-1}$, that is $c_i \log i \rightarrow 0$ as $i \rightarrow \infty$ (see Lindgren and H. Rootzén, 1987)¹. As it will be shown later in this paper, the long-run frequency of new records for $I(1)$ time series decreases at the slower rate of $n^{i-1/2}$:

Figures 1, 2, 3 and 4 show respectively the sequences of running ranges corresponding to a realization of a random walk process $y_t = y_{t-1} + e_t$; where $e_t \gg N(0; 1)$ (Figure 1), a stationary Gaussian AR(1) process $y_t = 0.5y_{t-1} + e_t$ (Figure 2), an AR(1) process $y_t = 0.5y_{t-1} + \varepsilon_t$ where the model errors ε_t follow a Student-t distribution with 5 degrees of freedom (Figure 3), and finally the same model with ε_t following a Cauchy distribution (Figure 4). We remark that for the random walk, the sequence of running ranges escalates indefinitely, whereas it does not in the other cases. The two latter figures show that thick-tailed error distributions or mere infinite variance do not imply the divergence of the running ranges. Such a divergence is rather caused by strong serial dependence.

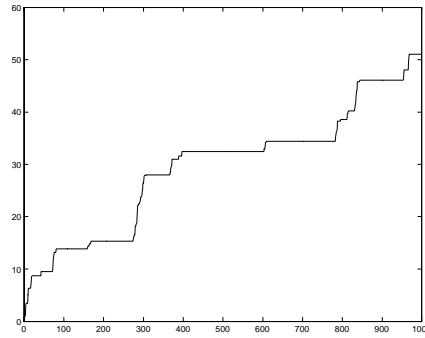


Figure 1. Sequence of running ranges for a Gaussian random walk process $y_t = y_{t-1} + e_t$; where $e_t \gg N(0; 1)$.

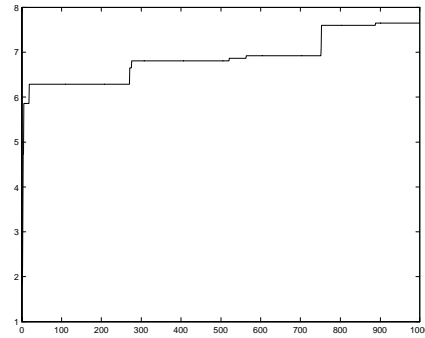


Figure 2. Sequence of running ranges for the Gaussian stationary AR(1) process $y_t = 0.5y_{t-1} + e_t$; with $e_t \gg N(0; 1)$.

¹ Any time series with exponentially decaying covariances satisfies the "Berman condition".

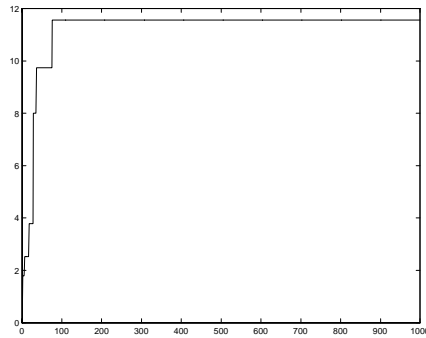


Figure 3. Sequence of running ranges for the AR(1) process $y_t = 0.5y_{t-1} + \varepsilon_t$; where ε_t is an i.i.d. sequence of random variables with a Student-t distribution with 5 degrees of freedom.

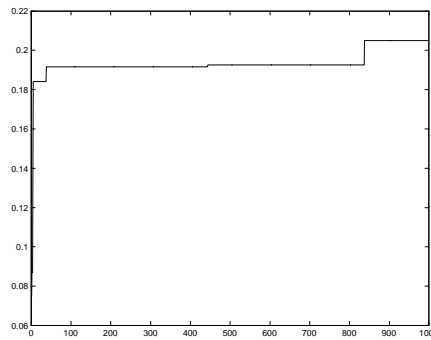


Figure 4. Sequence of running ranges for the AR(1) process $y_t = 0.5y_{t-1} + \varepsilon_t$; where ε_t is an i.i.d. sequence of random variables with a Cauchy distribution.

Figures 5 and 6 illustrate the same fact by showing an estimate of the long-run probability of new records: The probability was estimated from 1000 replications and for a sample size of $n = 1000$: Remark that when x_t is a stationary AR(1) series this probability estimate goes to zero for increasing t faster than when x_t is a random walk. Thus the presence of a long-wave pattern in the series seems to be related to a higher persistence of records.

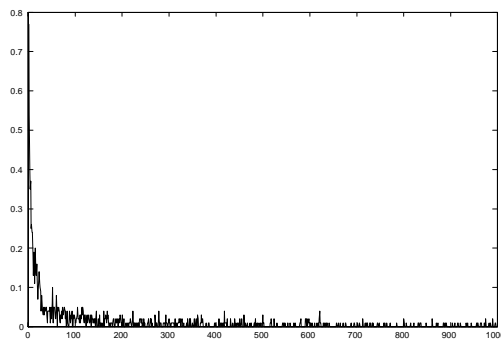


Figure 5. Estimate of $P \left(\Phi R_t^{(x)} > 0 \right)$ versus t when $x_t = 0.6x_{t-1} + \varepsilon_t$ and $\varepsilon_t \gg \text{Nid}(0; 1)$.

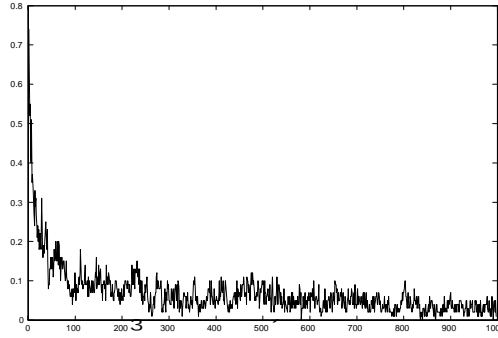


Figure 6. Estimate of $P \left(\Phi R_t^{(x)} > 0 \right)$ versus t when $x_t = x_{t-1} + \varepsilon_t$ and $\varepsilon_t \gg \text{Nid}(0; 1)$.

Another equivalent aspect of the previous property is captured by the mean interarrival times between consecutive records. Records show up as jumps in the sequence of running ranges. Figures 7 and 8 show respectively the mean interarrival times between the first 50 consecutive maxima for a random walk $y_t = 0.5y_{t-1} + e_t$, and for a stationary AR(1) time series generated by the model $y_t = 0.5y_{t-1} + e_t$; where $e_t \gg \text{Nid}(0; 1)$ in either case. The mean jump interarrival times were estimated from 1000 replications of the various models, each with a sample size of $n = 1000$. The figures clearly show how the sequence of interarrival times is stable for a random walk, while it explodes for the stationary AR(1) process. Therefore new records become increasingly rare for the latter, but do persist in time series with a unit root.

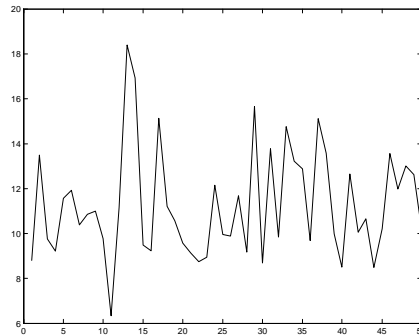


Figure 7. Mean interarrival times between the first 50 consecutive maxima of the Gaussian random walk process $y_t = y_{t-1} + e_t$; where $e_t \gg \text{Nid}(0; 1)$:

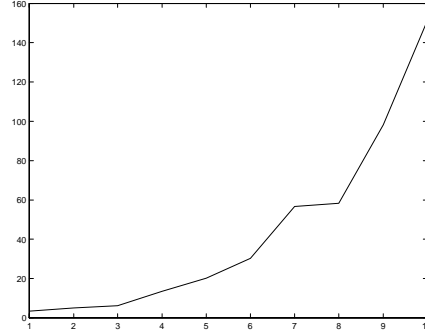


Figure 8. Mean interarrival times between the first 50 consecutive maxima of the Gaussian AR(1) process $y_t = 0.5y_{t-1} + e_t$; where $e_t \sim N(0, 1)$:

3 A Range Unit Root (RUR) Test

In this section we present the test statistic upon which the proposed unit root testing methodology is based. Then we analyse its small-sample behavior under the null hypothesis of a single unit root, provide some asymptotic results concerning this behavior, and finally, study its small-sample power performances against stationary AR(1) alternatives.

3.1 The test statistic

In the sequel we will consider the statistic $J^{(n)}$ defined below for testing the null hypothesis of a random walk $x_t = x_{t-1} + \varepsilon_t$ where the errors $\varepsilon_{t|t-1}$ are a sequence of i.i.d. random variables having zero mean and variance $\frac{1}{2}$. The corresponding testing device will be referred to as the Range Unit Root (RUR hereafter) test.

$$J^{(n)} = n^{-1/2} \sum_{t=1}^n 1(\Phi R_t^{(x)} > 0); \quad (1)$$

The statistic $J^{(n)}$ can be interpreted as a measure of the errors committed in predicting the range of x_t at t , $R_t^{(x)}$, by means of its value at time $t-1$, $R_{t-1}^{(x)}$. As the sample size approaches infinity, $J^{(n)}$ yields a global measure of the one-step-ahead range prediction errors, that is of the divergence rate of the range sequence for x_t . Notice also that $n^{-1/2} J^{(n)}$ represents the proportion of these prediction errors in a sample of size n ; while $n^{1/2} J^{(n)}$ is the number of records of the time series x_t up to time n :

Given the non-ergodic nature of x_t under the null hypothesis, the normalized number of records in the sample, $J^{(n)}$; does not converge to zero but to a

non-degenerate random variable, as it will be shown later. On the contrary, when $x_t \gg I(0)$, $J^{(n)}$ converges in probability to zero. This means that when $x_t \gg I(0)$, $R_{t-1}^{(x)}$ is a more efficient predictor of $R_t^{(x)}$ than when x_t contains a unit root. Consequently, the test statistic $J^{(n)}$ will be expected to take comparatively large values for $I(1)$ time series while small for $I(0)$ time series. Farther in the paper we will show also that this test is either robust or invariant to a number of departures from the null hypothesis.

3.2 Small-sample behavior under the null

Summary statistics for $J^{(n)}$ under the null hypothesis are given in Table 1 for a sample size of $n = 1000$, and for $Nid(0; 1)$ errors. The estimates were obtained from 10000 replications of the null model $x_t = x_{t-1} + \varepsilon_t$.

Summary Statistics	minimum	maximum	mean	median	std. dev.
Sample estimates	0.80	4.65	2.11	2.02	0.63

Table 1.

Table 2 shows estimates of the critical values of $J^{(n)}$ obtained from 10000 replications of the null model, and for eight different sample sizes and six significance levels ($\alpha = 0.01; 0.025; 0.05; 0.10; 0.90; 0.95$).

$\alpha \backslash n$	100	250	500	1000	2000	3000	4000	5000
0:01	0.9	0.9391	1.0119	1.0435	1.1180	1.1502	1.0594	1.0465
0:025	1.0	1.0752	1.1180	1.2333	1.1404	1.1502	1.2491	1.1031
0:05	1.1	1.2017	1.2075	1.2649	1.2746	1.3145	1.3123	1.3152
0:10	1.3	1.3282	1.3416	1.3598	1.2969	1.3510	1.3756	1.4425
0:90	2.8	2.9725	2.9963	3.0990	3.2870	2.9212	3.0042	2.7577
0:95	3.1	3.2888	3.3541	3.3520	3.6001	3.2498	3.3046	3.1396

Table 2.

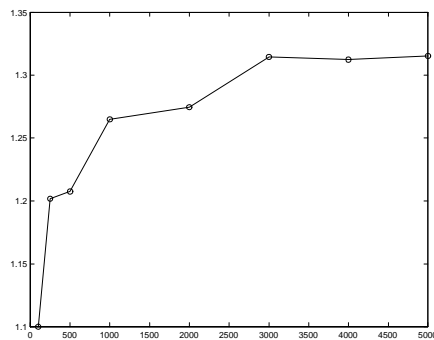


Figure 9. Convergence of the empirical critical values of $J_0^{(n)}$ as n increases.

Figure 9 shows the convergence of the 5%-level empirical critical values of $J^{(n)}$ (vertical axis) towards the asymptotic values with increasing sample size (horizontal axis), while Figure 10 shows the empirical density of $J^{(n)}$ estimated by kernel smoothing, again under the null hypothesis of a random walk with $Nid(0; 1)$ errors. The estimates were obtained from 1000 replications and for different sample sizes using the Epanechnikov kernel (see for example Silverman, 1986 and Hardle, 1990), which is optimal in the Mean-Square Error (MSE) sense².

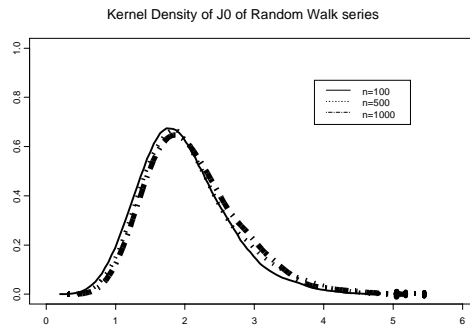


Figure 10. Plot of the empirical density of $J_0^{(n)}$ under the null hypothesis $H_0 : x_t = x_{t-1} + z_t$, where $z_t \gg Nid(0; 1)$. The density was estimated using the Epanechnikov kernel on different sample sizes.

4 Asymptotics

A basic result regarding the behavior of the records of a random walk is that the long-run frequency of these records is equal to zero. Proposition 1 formally establish this result, which is proved in the Appendix and will be used subsequently. This proposition relies on an already classical set of conditions for the model errors $\varepsilon_{i,j-1}$ proposed by Phillips (1987).

²

The density of $J_0^{(n)}$ was estimated as:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right);$$

with $K(\cdot)$ given by $K(u) = \frac{3}{4}(1 - u^2)1(|u| \leq 1)$:

$$C1. E(z_t) = 0:$$

$$C2. \sup_t E(j^2 t^p) < \infty \text{ for some } p > 2:$$

$$C3. 0 < \lim_{n \rightarrow \infty} E \left(\frac{1}{n} \sum_{i=1}^n z_i^2 \right)^2 < \infty:$$

$$C4. \{z_t\}_{t=1}^\infty \text{ is strong mixing with mixing coefficients } \alpha_m \text{ satisfying } \sum_{m=1}^\infty \alpha_m^{2/p} < \infty:$$

Assumption C2 allows the possibility of heteroskedastic errors by just requiring the finiteness of the higher order moments of the innovations. Condition C3 imposes bounds on the long-run variance of x_n under the null hypothesis. The lower bound rules out pathological cases such as $I(1)$ processes. Finally, assumption C4 allows trading an increasing degree of temporal dependence against a decreasing degree of heteroskedasticity (and viceversa) in the process.

Proposition 1 Let $x_t = x_{t-1} + z_t$ where $\{z_t\}_{t=1}^\infty$ satisfies assumptions C1-C4, and let $x_{t:t} = \max\{x_1, \dots, x_t\}$ and $x_{1:t} = \min\{x_1, \dots, x_t\}$. Then we have

$$\lim_{t \rightarrow \infty} P(x_t = x_{t:t}) = \lim_{t \rightarrow \infty} P(x_t = x_{1:t}) = 0: \quad (2)$$

Proof. In Appendix A1. ■

A most important question is the appropriate scaling needed for the sequence of partial sums $\sum_{t=1}^n (\Phi R_t^{(x)} > 0)$ in order to converge to a non-degenerate random variable under the null hypothesis H_0 , as well as the distribution of such a variable. Our main result establishes that under H_0 the normalized sequence of partial sums $J^{(n)} = n^{1/2} \sum_{t=1}^n (\Phi R_t^{(x)} > 0)$ converges weakly to a random variable whose pdf is the auto-convolution of the pdf of a Brownian local time at zero. Under the alternative hypothesis of stationarity, and mediating very mild restrictions on the degree of serial dependence of x_t , the sequence of partial sums $\sum_{t=1}^n (\Phi R_t^{(x)} > 0)$ diverges at a much lower rate, thus leading to $J^{(n)}$ as $n \rightarrow \infty$:

Definition 2 (Local Time of a Brownian Motion Process) (Lévy, 1948)
Let $B(\cdot)$ represent a Brownian motion process in \mathbb{R} , and let $L_B(x; t)$ be defined as

$$L_B(x; t) = \lim_{\pm \neq 0} \frac{1}{2\pm} \int_0^t 1[x - \pm \leq B(s) \leq x + \pm] ds \quad (3)$$

$L_B(x; t)$ is a continuous increasing process in x called the local time of B at x : It measures the amount of time the Brownian motion spends in the neighborhood of x . It can also be interpreted as the "spatial density" of the occupation time $\int_0^t 1[x - \pm \leq B(s) \leq x + \pm] ds$:

Theorem 3 Let $x_t = \sum_{i=1}^t z_i$ where $f_{i,j}^2$ are continuous i.i.d. random variables with bounded and symmetric pdf, zero mean and finite variance $\frac{1}{42}$. Suppose that x_0 has also a bounded pdf and finite variance. And let $J^{(n)} = J_1^{(n)} + J_2^{(n)}$ with $J_1^{(n)} = n^{i-1/2} \sum_{t=1}^n 1(x_t = x_{t,i})$ and $J_2^{(n)} = n^{i-1/2} \sum_{t=1}^n 1(x_t = x_{1,i})$. Then we have

1.

$$J_1^{(n)} \rightarrow \frac{E f_{1,j}^2}{\frac{1}{42}} I_B(0; 1) \quad (4)$$

$$J_2^{(n)} \rightarrow \frac{E f_{1,j}^2}{\frac{1}{42}} I_B(0; 1) \quad (5)$$

$$P(J_1^{(n)} < h) = \frac{2}{\sqrt{2\pi} \frac{E f_{1,j}^2}{\frac{1}{42}}} \int_0^h \exp\left(-\frac{u^2}{2 \frac{E f_{1,j}^2}{\frac{1}{42}}}\right) du; \quad h > 0; \quad i = 1, 2 \quad (6)$$

2.

$$J_1^{(1)} \text{ and } J_1^{(2)} \text{ are independent random variables} \quad (7)$$

and

$$P(J_1^{(1)} < h) = \frac{1}{\sqrt{2\pi} \frac{E f_{1,j}^2}{\frac{1}{42}}} \int_0^h \exp\left(-\frac{v^2 + 2}{4}\right) [1 - \Phi(v)] dv; \quad (8)$$

where $\Phi(\cdot)$ is the probability distribution function of a standard Normal random variable

3. If x_t is a stationary Gaussian series with covariance sequence $f_{c,i} = \text{Cov}(x_t, x_{t+i})g_i$ satisfying $c_i \log i \rightarrow 0$ as $n \rightarrow \infty$ (Berman condition): Then we have

$$J^{(n)} \rightarrow 0; \quad (9)$$

and thus the RUR test is consistent against this sort of alternatives.

Proof. In Appendix A2. ■

To obtain the asymptotic null distribution of $J^{(n)}$ we proceed by first splitting it into two terms:

$$\begin{aligned} J^{(n)} &= n^{i-1/2} \sum_{i=1}^n 1(\Phi R_i^{(x)} > 0) \\ &= n^{i-1/2} \sum_{i=1}^n 1(x_i - x_{i,i} = 0) + n^{i-1/2} \sum_{i=1}^n 1(x_i - x_{1,i} = 0) \end{aligned}$$

and then realizing that each term in this sum is the normalized number of visits to the origin of the two $I(1)$ processes with asymptotically i.i.d. innovations:

$$y_t = x_t - x_{t;t} \quad (10)$$

$$y_t^0 = x_t - x_{1;t} \quad (11)$$

Figures 11 and 12 show a realization of these processes obtained from the same random walk with i.i.d. Gaussian innovations with zero mean.

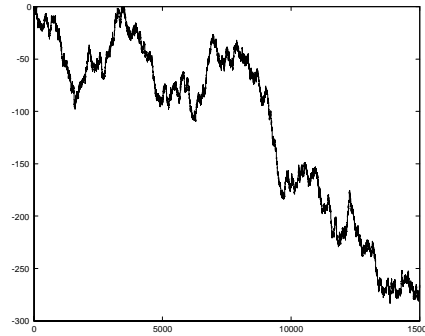


Figure 11. Realization of the process $y_t = x_t - x_{t;t}$; where x_t is a random walk with $N(0, 1)$ errors.

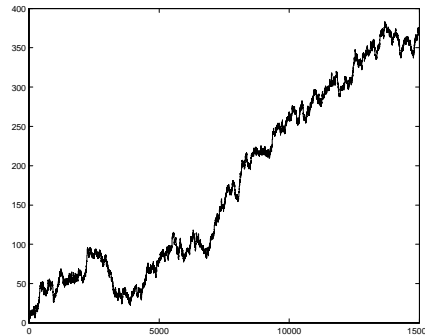


Figure 12. Realization of the process $y_t^0 = x_t - x_{1;t}$; where x_t is a random walk with $N(0, 1)$ errors.

The $I(1)$ nature of y_t and y_t^0 allows the application of a result by Burrige and Guerre (1996) for the asymptotic distribution of the normalized number of level crossings of a random walk, and which leads straightforwardly to ours. Quite interestingly, the asymptotic distribution of $J^{(n)}$ does not depend at all on the innovations' distribution (in particular of their variance, $\frac{1}{2}$), in spite that such a dependence exists for the asymptotic distribution of $J_i^{(n)}$: This dependence comes from the scaling factor $a = \sqrt{E f^2 J} = \frac{1}{2}$ which varies from one

error distribution to another. For example, if the innovations z_t are Gaussian then $a = \frac{E f_j^2 g}{2\sigma^2} = \frac{1}{4}$, and thereby even the asymptotic distribution of the normalized number of upper (or lower) records, $J_1^{(n)}$ ($J_2^{(n)}$); is unaffected by errors' variance, σ^2 . However, this case is rather exceptional since for all other common distribution the value of a is sensitive to its shape, or equivalently to the tails. This is shown below for some typical error distributions with shape parameter denoted by ρ .

Probability Distribution of Model Errors $f_{t g_{t-1}}$	$a = \frac{E f_j^2 g}{2\sigma^2}$
Student t with ρ degrees of freedom	$\frac{\rho-2}{4} \frac{\Gamma(\frac{\rho-1}{2})}{\Gamma(\frac{\rho}{2})}$
Log-Normal	$\frac{1}{\exp(\rho^2)} \frac{1}{\rho}$
Gamma	$\frac{\Gamma(\rho+1)}{\Gamma(\rho)} \frac{1}{\rho}$
Weibull	$\frac{\Gamma(\frac{\rho+1}{c})}{\Gamma(\frac{\rho+2}{c})} \frac{1}{\rho}$

Therefore, in general, the asymptotic distribution of the statistics $J_1^{(n)}$ and $J_2^{(n)}$ has different support depending on the shape of the model error distribution, which acts as a nuisance parameter. However, the asymptotic distribution of our test statistic $J^{(n)}$ is "error-model-free", and is therefore not affected by such parameters. This is contrast with the unit root testing device suggested by Burrige and Guerre (1996), based on the number of crossings, which in fact, not only does not improve DF's performances but depends on these nuisance parameters.

Another immediate consequence of the error-model-free property of the RUR test is that, contrary to the behavior of DF tests (see Hamori and Tokihisa, 1997), it is invariant to shifts in the model error variance.

To ensure the consistency of our test against general stationary alternatives we need to impose a restriction on the amount of serial dependence of the process. The following condition is similar in spirit (although much weaker) to the strong-mixing condition and allows borrowing results from the asymptotic theory of records for i.i.d. processes.

Condition 4 Let $\{x_t|g_{t-1}\}$ be a stationary sequence of random variables with

$$F_{i_1, \dots, i_n}(u_1; \dots; u_n) = P\{x_{i_1} \leq u_1; \dots; x_{i_n} \leq u_n\}$$

representing its finite-dimensional distribution function. Write $F_{i_1, \dots, i_n}(u) = F_{i_1, \dots, i_n}(u; \dots; u)$ for economy of notation and define

$$\theta_{n,l}^a = \max_{1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n} |F_{i_1, \dots, i_p, j_1, \dots, j_q}(u) - F_{i_1, \dots, i_p}(u) F_{j_1, \dots, j_q}(u)|$$

with $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$; $j_1 \leq i_p$. The sequence $\{x_t|g_{t-1}\}$ is said to satisfy condition $D(u_n)$ if there exists a sequence of numbers $l_n = o(n)$ such that $\theta_{n,l_n}^a \rightarrow 0$.

Among the processes that satisfy the $D(u_n)$ are the Gaussian processes satisfying the so-called "Berman condition". If x_t is one such process then the joint distribution of any fixed set of extreme statistics converges to the same limit as if the variables were i.i.d: (Lindgren and Rootzén, 1987). As a consequence, we must expect $n^{1/2} J^{(n)} \gg O(\log n)$, or equivalently, $J^{(n)} \gg O(n^{1/2} \log(n))$ as $n \rightarrow \infty$, and the consistency against this class of alternatives is proved: Remark that the Berman condition is not very demanding, since it is satisfied by any process with exponentially decaying covariance function, among which are all the stationary Gaussian ARMA processes. However, the Gaussian condition seems to be too restrictive, since in our simulations the power of the RUR test did not seem to vary significantly on stationary alternatives with different error distributions (such as Cauchy's and the Student t).

The RUR test is also asymptotically immune to the presence of additive noise in the series. The proposition below states formally this property. We will use the following definition of an $I(0)$ process proposed by Davidson (1998).

Definition 5 A time series $\{s_t\}_t$ is $I(0)$ if the process S_n defined on the unit interval by

$$S_n(r) = \frac{1}{n^{1/2}} \sum_{t=1}^{[nr]} s_t - \frac{1}{n^{1/2}} [nr] E(s_t); \quad 0 \leq r \leq 1; \quad (12)$$

where $\frac{1}{n^{1/2}} = \text{Var} \left(\sum_{t=1}^n s_t \right)^{-1/2}$, verifies

$$S_n(r) \Rightarrow B(r)$$

with $B(r)$ representing a standard Brownian motion on $[0,1]$.

Proposition 6 Let x_t be defined as $x_t = \sum_{i=1}^t z_i + s_t$ where $\{z_i\}_{i=1}^t$ is a family of i.i.d: random variables with zero mean and finite variance $\frac{1}{2}$, and with s_t representing an $I(0)$ time series process in the sense of the previous definition. Then

$$P(fJ^{(1)}) < hg = \frac{1}{2^{1/2}} \int_0^{\infty} \exp\left(-i \frac{v^2+2}{4}\right) [1 - \exp(-v)] dv; \quad (13)$$

Proof. See Appendix A3. ■

Figure 13 illustrates the noise immunity property of the RUR test by giving the power curves as a function of the noise variance (horizontal axis), $\frac{1}{n^{1/2}}$. For

any finite value of $\frac{3}{4}\sigma^2$ the power curves tend to collapse at the nominal size of the test (here 5%) with increasing sample size. The same power curves are plotted in Figure 14 for the DF test. These figures suggest that for any finite sample size there is a threshold value for the noise variance beyond which the size distortion in the DF case reaches its maximum. This is contrast with the RUR case for which the size distortion increases very slowly with increasing noise variance while approaching its nominal value with increasing sample size. Consequently, in finite samples, the amount of size distortion is remarkably larger for the DF test than for the RUR test.

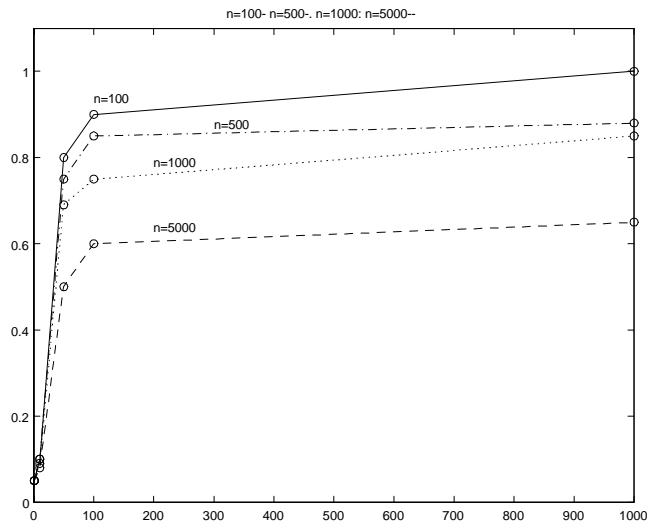


Figure 13. Power curves of RUR test as a function of the noise variance, for different sample sizes.

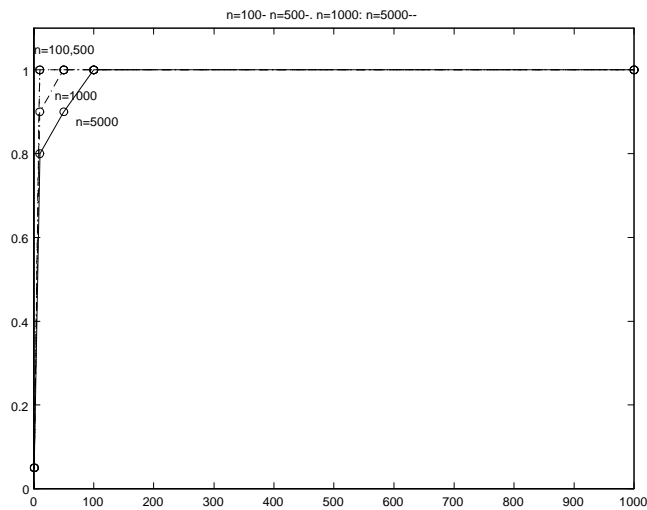


Figure 14. Power curves of DF test as a function of the noise variance, for different sample sizes.

If we let $\frac{1}{2}_{df}$ denote the DF test statistic and let $\frac{1}{2}_{\alpha}$; J_{α} represent the critical values of the DF and RUR tests, respectively, at the significance level α ; then the gain $G(n; \frac{1}{4}_S^2)$ in terms of size distortion offered by the RUR test for a sample size equal to n and a noise variance equal to $\frac{1}{4}_S^2$ could be measured by the ratio of probabilities:

$$G(n; \frac{1}{4}_S^2) = \frac{1 - P\left(\frac{1}{2}_{df}^{(n)} < J_{\alpha}\right)}{1 - P\left(\frac{1}{2}_{df}^{(n)} < \frac{1}{2}_{\alpha}\right)} = \frac{P\left(\sum_{i=1}^n z_i^2 + s_t; \text{var}(s_t) = \frac{1}{4}_S^2\right)}{P\left(\sum_{i=1}^n z_i^2 + s_t; \text{var}(s_t) = \frac{1}{4}_S^2\right)}.$$

The aforementioned property of the RUR test statistic could then be expressed as:

$$\lim_{\frac{1}{4}_S^2 \rightarrow 1; n \rightarrow \infty} G(n; \frac{1}{4}_S^2) = 1 :$$

Figure 15 shows how the gain in noise immunity offered by the RUR test over the DF test increases with both increasing sample size and increasing variance of the additive noise component.

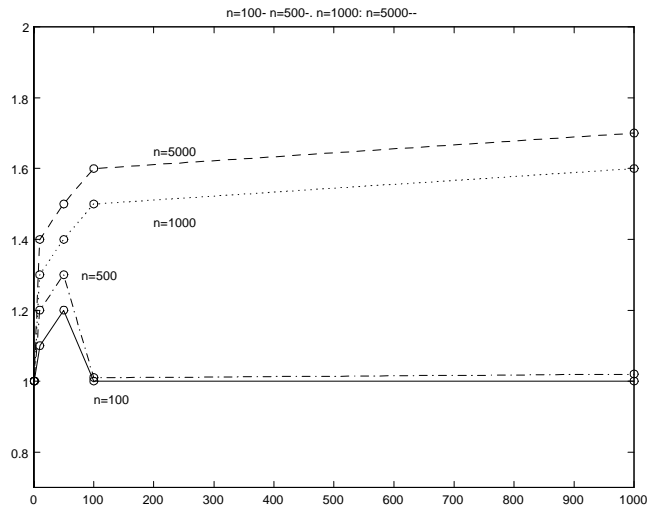


Figure 15. Gain in noise immunity of the RUR test over the DF test as a function of the noise variance and for different sample sizes

5 Power Performances and Consistency of the Test

In this section we investigate the power performances of the RUR test and its consistency against stationary, trending and integrated alternatives. First of all, it is easy to show that the test is consistent against stationary alternatives. To show it, recall from Section 2 that for such alternatives we can expect the sequence of ranges to behave similarly as if x_t was an $i:i:d$ sequence, that is:

$$\sum_{i=1}^n 1(\Phi R_i^{(x)} > 0) = O(n^{i-1} \log n)$$

from where

$$J^{(n)} = O(n^{i-1} \log n):$$

Therefore the test is consistent since $n^{i-1} \log n \rightarrow 0$ as $n \rightarrow \infty$; while $P(J^{(1)} = 0 | H_0) = 0$: A similar behavior applies on $I(i, k)$ time series with $k > 0$ since the degree of mean reversion is even more pronounced in this case. The following simple device will allow to discriminate between the stationary and the integrated case. Let B denote the delay operator and let $x_t^{(0)} \triangleq x_t$: Note that if $x_t \gg I(0)$ then the time series defined by $x_t^{(1)} \triangleq \sum_{j=0}^{t-1} B^j x_{t-j}$ is $I(1)$: Similarly, if $x_t \gg I(i, k)$ with $k > 0$ then $k+1$ will be the smallest positive integer such that $x_t^{(k+1)} \gg I(1)$; or equivalently, such that $J^{(n)}$ does not vanish asymptotically. By mere inversion of the argument, if k is the smallest nonnegative integer such that the null hypothesis is not rejected on $x_t^{(k)}$ then x_t will very likely be $I(i, k)$:

The small-sample power of the test against stationary AR(1) alternatives is shown in Table 3 below using the estimated critical values at the 5% significance level, and from 10000 replications of the alternative model $x_t = b x_{t-1} + \varepsilon_t$, with $\varepsilon_t \gg N(0, 1)$. The autoregressive parameter b was allowed to take different values (0.5; 0.8; 0.9) as well as the sample size n (100; 250; 500). The DF performances appear in brackets.

$n \setminus b$	0.5	0.8	0.9	0.95	0.99
100	0.8 (1)	0.6 (0.99)	0.5 (0.5)	0.4 (0.18)	0.12 (0.0375)
250	1 (1)	1 (1)	1 (1)	0.8 (0.7)	0.47 (0.0760)
500	1 (1)	1 (1)	1 (1)	1 (0.99)	0.72 (0.39)

Table 3.

These results show that the DF test outperforms the RUR test in only two cases: (i) when the sample size is comparatively small ($n = 100$), and (ii) when the autoregression parameter b is not close to the nonstationary border. On near-unit root stationary time series, however, the RUR test outperforms

remarkably DF test. This is not surprising knowing that unit root detection procedures based on prediction errors have nearly optimal properties in the vicinity of the null (see Sanchez, 2003).

The power curves are plotted in Figure 16 for three different sample sizes ($n = 100; 250; 500$). The continuous and dotted lines correspond respectively to the DF and the RUR test results.

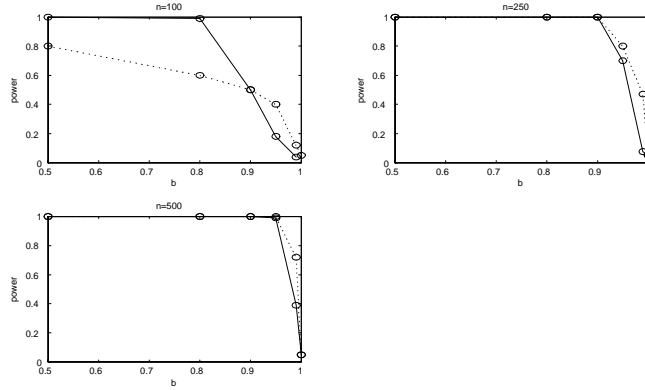


Figure 16. Plots of power estimates of the DF and RUR tests against stationary Gaussian AR(1) series.

Therefore as compared to the DF test, the RUR test establishes a sharper border between the null hypothesis of unit root and the stationary AR(1) alternatives. This can be explained by the invariance of the RUR test statistic $J^{(n)}$ with respect to the infinite variance \mathcal{H}_X^2 of the stationary alternative, and which follows trivially from the fact that

$$1(\Phi R_t^{(x)} > 0) = 1(\mathcal{H}_X^{-1} \Phi R_t^{(x)} > 0):$$

On trending alternatives, the RUR test is also consistent. To see this, we invoke a classical result by Feller (1971) which states that on random walks with nonzero drift, that is when $\mu_2 = E(\epsilon_t) \neq 0$; the renewal counting process of records $N(t) = \sum_{i=1}^t 1(\Phi R_i^{(x)} > 0)$ satisfies:

$$\lim_{t \rightarrow \infty} t^{-1} N(t) = O(1):$$

As a consequence, $J^{(n)} = O(n^{1/2}) \rightarrow \infty$ as $n \rightarrow \infty$ under such alternatives.

A similar divergent behavior of the RUR test statistic occurs when $x_t \gg I(k)$ with integration order $k > 1$; or when x_t is a stationary time series fluctuating around a deterministic trend. To distinguish between these two cases consider the following time series models:

- a) $x_t = x_{t-1} + \epsilon_t$ with $E(\epsilon_t) = \mu_2 \neq 0$;
- b) $x_t = y_t + \mu t$ where $y_t \gg I(0)$:

Notice that under model a) $\Phi x_t \gg I(0)$; while under model b) $\Phi x_t \gg I(1)$: So discrimination between model a) and b) is reduced to a previously solved problem.

6 The RUR Test Statistic under Departures from the Standard Assumptions

Another important property of the RUR test is its robustness to departures from the standard assumptions. In this paper, we consider three types of departures: a) when a stationary time series undergoes structural breaks; b) when $I(1)$ time series are corrupted by additive outliers; and c) when $I(1)$ time series are nonlinearly transformed. In the sequel we study the small sample behavior of the RUR test in the presence of each of the above-mentioned departures from the standard unit-root tests assumptions.

6.1 Stationary time series with level shifts

Many economic and financial time series such as inflation, nominal and real interest rates can be trend-stationary with a structural break in the unconditional mean which affects the standard inferential procedures and often makes constant coefficient models to perform poorly in practice (see for instance Perron, 1990, and Malliaropoulos, 2000). The literature on testing for unit roots in the presence of both known and unknown break points is large (see Maddala and Kim, 1998 for a review). Perron (1989), Vogelsang (1990) and Perron and Vogelsang (1992) reported evidence that structural breaks can make an $I(0)$ time series behave locally as $I(1)$ and, as a result, these breaks are able to fool standard unit root tests (this is shown here by means of simulations). More precisely, Perron (1989) and Rappoport and Reichlin (1989) showed via Monte Carlo experiments that time series such as GNP previously modelled as $I(1)$, appear as $I(0)$ if we allow for a segmented trend in the model during the oil crisis. In brief, if the permanent break is not explicitly taken into account standard unit root tests tend to find too many unit roots. This drawback of standard DF tests is usually referred to in the literature as the "Perron phenomenon". Leybourne, Mills and Newbold (1998) reported that a "converse Perron phenomenon" is also often observed, which consists in the spurious rejection of the null hypothesis of a unit root when an early break appears in an $I(1)$ time series. DF test outcomes can therefore be very misleading. Moreover, the critical values of standard unit root tests depend on the new unknown nuisance parameters such as the number of breaks and their timing, which has led several authors (see Zivot and Andrews, 1992; Perron and Vogelsang, 1992; Banerjee, Lumsdaine and Stock, 1992; and Stock, 1994) to propose complex recursive and sequential testing procedures in order to estimate these parameters. In the light of all these difficulties, the RUR test combines simplicity with robustness and enhanced power performances. In

the sequel we analyse these performances when confronted to the alternative of a stationary AR(1) time series corrupted by one or two structural breaks. First we will consider the case of a single structural break in the series in the middle of the sample. The break is modeled as a dummy variable defined by $D_t = 0$ for $t \leq n/2$ and $D_t = 1$ for $t > n/2$. Specifically, we consider the alternative model is $x_t = 0.5 x_{t-1} + s D_t + \varepsilon_t$. Table 4 provides power estimates at the 5% significance level from 10000 replications for different values of the sample size n and of the local break size s . The power performances of the Dickey-Fuller test appear in brackets.

$n \setminus s$	4	8	12
100	0.2 (0.00)	0.08 (0.00)	0.07 (0.00)
250	0.7 (0.00)	0.6 (0.00)	0.6 (0.00)
500	1 (0.86)	1 (0.00)	1 (0.00)

Table 4.

We remark that except for the case of $s = 4$ and $n = 500$, the Dickey-Fuller (DF) test has a much stronger bias towards nonstationarity, and thus the RUR test is less prone to misinterpret structural breaks as permanent stochastic disturbances.

In a scenario allowing for multiple breaks, we should expect an even larger decrease in power for both the RUR and the DF tests. In order to assess these power losses, we performed another experiment which included two breaks at different locations in time. The alternative model was now $x_t = 0.5 x_{t-1} + s_1 D_{t;1} + s_2 D_{t;2} + \varepsilon_t$ with $D_{t;i}$ ($i = 1; 2$) representing dummy variables defined by $D_{t;i} = 0$ for $t \leq in/4$ and $D_{t;i} = 1$ for $in/4 < t \leq in/2$. Table 5 shows the power results at the 5% significance level obtained from 10000 replications of this model, for both the RUR and the DF tests (the DF figures given in brackets). Here $s_{1;2} = (s_1; s_2)^0$. The power estimates are given for different values of the sample size n (100,250,500), and of the break magnitudes s_1 and s_2 ($s_1 = 2; 4; 8$; and $s_2 = 4; 8; 12$; respectively). Once again, the RUR test outperforms the DF results in all cases, and is still remarkably powerful for the sample size $n = 500$, as far as the break size is not too large.

$n \setminus s_{1;2}$	(2,4)	(4,8)	(8,12)
100	0.07 (0.000)	0.005 (0.000)	0.000 (0.000)
250	0.5 (0.000)	0.200 (0.000)	0.05 (0.000)
500	1 (0.453)	0.7 (0.000)	0.6 (0.000)

Table 5.

To explain theoretically this robustness of the RUR test, consider the following AR(1) models, in one of which we allow for the possibility of a single

break through the innovations' dynamics:

- a) $x_t = ax_{t-1} + \varepsilon_t$; $\text{con } E(\varepsilon_t) = 0$; $|a| < 1$;
b) $x_t = ax_{t-1} + \varepsilon_t$; $\text{con } E(\varepsilon_t) = s1(t = t_0)$; $|a| < 1$;

Let $J_1^{(n)}$; $J_2^{(n)}$ be the RUR test statistics associated with the processes in model a) and b), respectively. Now if $|a| \gg 0$ and $t_0 \gg 0$ we will have

$$R_{t_0} = a + R_{t_0-1};$$

since on $I(0)$ processes $P(\Phi R_t > 0) = O(t^{-1}) \rightarrow 0$ for t large enough, that is $\Phi R_{t_0} = a$ with probability close to one. As a result, $J_2^{(n)} \rightarrow J_1^{(n)} + n^{1-2}$ for both n and t_0 large enough. But then $J_2^{(n)} \rightarrow J_1^{(n)} \neq 0$ as $n \rightarrow \infty$; which means that the consistency of the test is not affected by the presence of a level break, whatever the size of such break.

When several level breaks are involved, say m breaks, we can write $E(\varepsilon_t) = \sum_{i=1}^m s_i 1(t = t_i)$. Now suppose $t^* \equiv \min_{1 \leq i \leq m} t_i \gg 0$ and $0 < s^* \equiv \max_{1 \leq i \leq m} |s_i| < 1$; such that $P(\Phi R_{t^*} > 0) = O(t^{*-1}) \rightarrow 0$ and thereby $J_2^{(n)} \rightarrow J_1^{(n)} + n^{1-2} \sum_{i=1}^m s_i$ for large enough n . Therefore $J_2^{(n)} \rightarrow J_1^{(n)} \neq 0$ as $n \rightarrow \infty$; for finite m . What is more, the number of level breaks, m , can even grow indefinitely as $o(n^{1-2})$ without affecting the consistency of the test.

A similar argument can be used to show that the asymptotic size of the RUR test is unaltered by the presence of as much as $m = o(n^{1-2})$ level breaks superimposed on a $I(1)$ time series. Indeed, suppose we have the following two models:

- a) $x_t = x_{t-1} + \varepsilon_t$; $\text{con } E(\varepsilon_t) = 0$;
b) $x_t = x_{t-1} + \varepsilon_t$; $\text{con } E(\varepsilon_t) = \sum_{i=1}^m s_i 1(t = t_i)$;

with $t^* \gg 0$ and $0 < s^* < 1$; such that $P(\Phi R_{t^*} > 0) = O(t^{*-1-2}) \rightarrow 0$ and thereby $J_2^{(n)} \rightarrow J_1^{(n)} + n^{1-2} \sum_{i=1}^m s_i$. Then as far as $m = o(n^{1-2})$ we will get $J_2^{(n)} \rightarrow J_1^{(n)} \neq 0$ as $n \rightarrow \infty$; and the asymptotic size will be the same as in model a).

Notice that if, on the contrary, m is allowed to be $O(n^{1-2+\delta})$ with $\delta > 0$; $J_2^{(n)}$ will behave as if x_t had a trend, that is, $J_2^{(n)} \rightarrow 1$ as $n \rightarrow \infty$: Level breaks will then shift indefinitely the null distribution to the left leading to the rejection of the null hypothesis of an $I(1)$ time series.

6.2 Nonlinearly transformed I(1) time series

In practice, it is difficult and even sometimes impossible to know whether a time series exhibiting unit-root-like behavior is really I(1), or rather a monotonically nonlinear transformation of an I(1) series. With standard unit-root tests, misspecification of the true time series model may affect the rate of divergence of the test statistic, making it to behave inconsistently. The invariance to such nonlinearities would be therefore a desirable property of a unit-root test.

Granger and Hallman (1991) looked at the autocorrelation function of several nonlinear transformations of the original series and proposed a test invariant to monotonic transformations based on ranks. Ermini and Granger (1993) worked with the Hermite polynomial expansion of different nonlinear transformations of random walks, possibly with drift, and showed that the autocorrelation function is not always a reliable indicator of the degree of memory of nonlinear time series.

In what follows we analyse the small sample behavior of the RUR test in the face of several nonlinear transformations of random walks, and show that it is invariant to monotonic transformations. Table 6 shows the size estimated at the 5% significance level from 10000 replications of the different models and for $n = 100$; 250; and 500.

Transformation	100	250	500
1) x_t^2	0.079 (0.397)	0.170 (0.406)	0.178 (0.420)
2) x_t^2 , with $x_t > 0$; $8t$	0.03 (0.397)	0.059 (0.406)	0.048 (0.420)
3) x_t^3	0.038 (0.456)	0.057 (0.532)	0.049 (0.533)
4) $\exp(x_t)$	0.03 (0.92)	0.05 (1)	0.0469 (1)
5) $\exp(\frac{x_t}{75})$	0.054 (0.271)	0.0526 (0.271)	0.05 (0.301)
6) $\log(x_t + 100)$	0.043 (0.275)	0.064 (0.331)	0.051 (0.354)
7) $\log(\frac{x_t + \frac{P-T}{4}}{1 + \frac{P-T}{4}}); \frac{x_t + \frac{P-T}{4}}{1 + \frac{P-T}{4}} \in (0; 1)$	0.072 (0.347)	0.054 (0.349)	0.051 (0.354)
8) $\sin(x_t)$	0.8828 (1)	0.9986 (1)	1 (1)

Table 6

It can be observed that the size tends towards its correct value in all cases except when the transformation is non-monotonic (case 1), and when it stationarizes the series (case 8). To study more precisely the effect of the logarithmic nonlinearity, in case 7), we forced the variable to take most of its values in the interval (0; 1). This was done by transforming linearly the series prior to applying the logarithmic transformation. Since in this interval the function is not so well approximated by a straight line, one would expect a more noticeable size distortion, at least for the smaller sample size of $n = 100$. Overall, however, all the empirical sizes for the purely monotonic transformations seem to converge to the nominal size of 0.05 as the sample size grows. The invariance of $J^{(n)}$ to monotonic nonlinear transformations $g(\cdot)$ applied to the series x_t follows

immediately from the relations:

$$\begin{aligned} 1(g(x_t) > g(x_{t_i-1;t_i-1})) &= 1(x_t > x_{t_i-1;t_i-1}) \\ 1(g(x_t) < g(x_{t_i-1;t_i-1})) &= 1(x_t < x_{t_i-1;t_i-1}): \end{aligned}$$

Notice that such invariance holds not only under the null hypothesis but also under any alternative. This result is in fact related to the invariance of the number of level crossings in a series (in this case, the first differences of the sequence of running ranges) to monotonic transformations.

6.3 Integrated time series corrupted by additive outliers

Outlying observations is another source of problems for the time series analyst. These may occur for different reasons, ranging from measurement errors to recordings of unusual events such as wars, disasters and dramatic policy changes. Some commonplace outlier-inducing events in economic time series are union strikes, hoarding consumer behavior in response to a policy announcement, and computer breakdown effects on unemployment or sales data collection and processing, to name a few. Outliers can also appear as a result of misspecified estimated relationships or omitted variables (see for instance Peña, 2001).

Outliers are usually classified into two groups: Additive Outliers (AO) and Innovation Outliers (IO), of which the former ones have the most insidious effects on classical inference. In both cases, standard unit root tests are biased towards the rejection of the unit root hypothesis. An IO (AO) corresponds to an external error or exogenous change in the observed value of the time series at a particular instant, with (no) effect on the subsequent observations in the series. In the case of an AO at time T , instead of observing the original series x_t , we observe a corrupted series y_t , given by:

$$y_t = \begin{cases} x_t & t \neq T \\ x_t + s & t = T \end{cases}$$

where s represents the outlier magnitude.

There is a sort of duality between the effects of AO's and those of structural breaks on time series. Indeed while $I(0)$ time series subject to level shifts could be misinterpreted as $I(1)$, $I(1)$ time series corrupted by AO's might look like $I(0)$ provided that the outliers are sufficiently frequent and important in magnitude. In particular, it is known that the presence of AO's leads to a downward bias of the OLS parameter estimates in a stationary AR(1) process (Bustos and Yohai, 1986; Martin and Yohai, 1986) and thereby the DF test will have an actual size in excess of the nominal size, thus rejecting the unit-root hypothesis too often. The size distortion of the DF test in the presence of AO's was quantified by Franses and Haldrup (1994), who also demonstrated that the distribution of the autoregression parameter estimates changes dramatically when both the magnitude of the outliers and their frequency become large.

Traditionally, the presence of AO's has been dealt with either by attaching less weight to the extreme observations in the sample (damping) or by removing them with the inclusion of a dummy variable in the model (...ltering). A recent example of the "...ltering approach" can be seen in Arranz and Escribano (1998b). Single-stage (without prior damping or ...ltering of the outliers) robust unit-root tests were ...rst proposed by Lucas (1995a,b) and by Franses and Lucas (1997) using M-estimators with high breakdown point and e¢ciency, instead of OLS estimators. However, these tests were really conceived for dealing with fat-tailed distributions of the model errors and therefore were less powerful than standard unit root tests on Normally distributed errors. Alternatively, some authors have followed a likelihood-based approach where inference is made about a particular fat-tailed distribution rather than on the Gaussian distribution (Hoek, Lucas and Van Dijk, 1995; Rothenberg and Stock, 1997). The use of semiparametric and nonparametric statistics is another avenue of research in robust unit root testing. Hasan and Kroenker (1997) applied rank-based methods to this problem and reported improved power performances on time series corrupted by a few large observations. The RUR procedure also falls into the nonparametric category.

The results in Tables 7 and 8 show that the size distortions caused by the presence of an AO in the middle of the series and beyond are significantly smaller for the RUR test than for the DF test (shown in brackets). Our alternative hypothesis was now represented by the model $y_t = x_t + s_{t;\zeta}$ where $x_t = x_{t-1} + \varepsilon_t$, ζ denotes an integer no larger than the sample size, and $s_{t;\zeta}$ is a dummy variable defined by $s_{t;\zeta} = 1$ if $t = \zeta$ and zero elsewhere. The sizes were estimated at the 5% significance level, for different values of both ζ ($\zeta = n/25; n/10; n/5$) and the sample size n (100; 250; 500). It can be seen that when the AO appears near the end of the series (Table 8) the RUR test have even lower than nominal sizes.

$n \setminus \zeta$	$n=2$	$n=2 + 1$	$n=2 + 2$
100	0.0826 (0.2978)	0.0830 (0.2964)	0.0812 (0.2958)
250	0.0800 (0.1682)	0.0800 (0.1688)	0.0798 (0.1670)
500	0.0644 (0.1130)	0.0640 (0.1102)	0.0642 (0.1096)

Table 7.

$n \setminus \zeta$	$n \setminus n=20$	$n \setminus n=10$	$n \setminus n=5$
100	0.0212 (0.2964)	0.0244 (0.2990)	0.0352 (0.2980)
250	0.0392 (0.1704)	0.0422 (0.1660)	0.0484 (0.1656)
500	0.0446 (0.1106)	0.0472 (0.1104)	0.0510 (0.1118)

Table 8.

Unfortunately, an early AO will produce a jump in the sequence of ranges which may prevent other jumps from being counted by the RUR test statistic,

thus biasing our test towards the rejection of the null hypothesis of unit root. The bias will be larger the sooner the outlier appears in the series. In order to grasp more closely this problem, we performed another Monte Carlo experiments in which a single AO is introduced near the origin (Table 9).

$n \setminus j_0$	$n=25$	$n=10$	$n=5$
100	0.3778 (0.2956)	0.3192 (0.2964)	0.2432 (0.3002)
250	0.2746 (0.1672)	0.2230 (0.1668)	0.1700 (0.1676)
500	0.1930 (0.1114)	0.1588 (0.1112)	0.1188 (0.1110)

Table 9.

The results show that when the AO appears within the first quarter of the sample, the RUR test seems to offer no real improvement over the DF test.

To give a flavor of what is going on in this case, suppose we have an AO early in the series at time $t = t_0$; and suppose that its magnitude, s , is such that $\Phi R_{t_0}^{(x)} = \max_{1 \leq t \leq n} \Phi R_t^{(x)}$. Such a large outlier will prevent new records from occurring at $t > t_0$; and therefore $\Phi R_t^{(x)} = 0$ for $t > t_0$. It follows that

$$J^{(n)} = n^{1-\alpha} \sum_{i=1}^{\lfloor n^\alpha \rfloor} 1(\Phi R_i^{(x)} > 0) \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty;$$

and the test will then be likely to reject the null hypothesis. Notice that the previous result still holds when the AO's location is allowed to increase with the sample size as fast as $O(n^{1-\alpha})$ with $\alpha > 0$:

Obviously, when more than one early AO appear the record count will be determined by the largest AO's location, but the real size of the test will grow to one, in the same way, as $n \rightarrow \infty$:

The relatively large size distortion of the RUR test in the presence of early AO's can be solved, however, by slightly modifying the test statistic so as to count also the records appearing when the series is observed in reversed order. The discussion on this variant of the test will be postponed for lack of space.

7 Concluding Remarks

Standard unit root tests suffer from a number of drawbacks when the usual assumptions are no longer justified. Apart from having low power on stationary near-unit root time series, they are also seriously affected by other aspects of real data such as parameter shifts, outliers and neglected nonlinearities.

In 1996 Burrridge and Guerre proposed a nonparametric unit root testing device based on the number of crossings. This test was sensitive to the tails of the error distribution and was inferior in other respects to DF's. In the light of these poor results, these authors concluded that there is no necessarily a gain

in robustness from using a nonparametric unit root test instead of the standard tests. In this paper, we have shown that there is no reason for such a pessimistic view about the possibilities of nonparametric methods. In particular, we have presented an nonparametric testing device, called the Range Unit Root (RUR) Test, which is robust to structural breaks either in the mean or in the variance, as well as to the presence of non-early additive outliers. The new method is also invariant to monotonic nonlinearities in the DGP and to the shape of the innovations' distribution, and outperforms the DF test in terms of power on stationary near-unit root alternatives. Finally, it is asymptotically immune to the presence of additive noise superimposed on an unobserved variable.

A. Appendix

In this section we provide the proofs for the theoretical results presented in previous sections. For this we need to invoke the following lemmas.

A0. Preliminary lemmas

Lemma 7 (Herrndorf's Invariance Principle). Let $\{f_t\}_{t=1}^{\infty}$ be a random sequence satisfying assumptions C1-C4, then defining

$$x_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} z_t \Rightarrow B(r)$$

where $B(\cdot)$ is a Brownian motion process on the interval $[0,1]$, $\frac{1}{\sqrt{n}}$ represents the long-run variance defined in C3, and " \Rightarrow " denotes convergence in distribution as $n \rightarrow \infty$:

Proof. See Herrndorf (1984). ■

Lemma 8 (Continuous Mapping Theorem). Let T be continuous function (except possibly on a set with Lebesgue measure equal to zero) such that: $T : C[0;1] \rightarrow C[0;1]$; where $C[0;1]$ denotes the space of cadlag functions on the interval $[0;1]$; Let $x_n(r)$ defined as in Lemma 9. Then

$$T(x_n(r)) \Rightarrow T(B(r)):$$

Proof. See Billingsley (1968). ■

Lemma 9 Let $x_n(r)$ defined as in Lemma 9. Under the assumptions C1-C4 we have:

$$\begin{aligned} x_n(r) &\Rightarrow \max_{s \in [0;1]} f x_n(s) = T_1(x_n(r)) \Rightarrow B(r) \Rightarrow \sup_{s \in [0;1]} f B(s) \\ x_n(r) &\Rightarrow \min_{s \in [0;1]} f x_n(s) = T_2(x_n(r)) \Rightarrow B(r) \Rightarrow \inf_{s \in [0;1]} f B(s): \end{aligned}$$

Proof. The proof follows from the CMT (Lemma 8) and the continuity of the functions T_1 and T_2 : ■

Lemma 10 (Lévy, 1948). Let $\{B(r)\}_{r \in [0;1]}$ represent a Brownian motion process on the interval $[0,1]$, and let $\mathcal{B}_1(r) = B(r) \Rightarrow \sup_{s \in [0;1]} f B(s)$ and $\mathcal{B}_2(r) = B(r) \Rightarrow \inf_{s \in [0;1]} f B(s)$: The processes $\{B(r)\}_{r \in [0;1]}$, $\{\mathcal{B}_1(r)\}_{r \in [0;1]}$ and $\{\mathcal{B}_2(r)\}_{r \in [0;1]}$ have the same probability distribution.

Proof. See Karatzas and Shreve (1988). ■

Lemma 11 (Lévy, 1948). Let $B_1(r) = B(r) - \sup_{s \in [0,1]} B(s)$ where $B(r)$ is a Brownian motion process in $[0,1]$. The random variables $B_1(r)$ and $I_B(0; r)$ have, for each value of $r \in [0, 1]$ the same probability distribution given by

$$P(B_1(r) \leq -b) = \frac{r}{2} \exp\left(-\frac{b^2}{2r}\right); \quad b > 0$$

Proof. See Karatzas and Shreve (1988). ■

Lemma 12 Let $x_t = x_{t-1} + z_t$ where z_t are i.i.d. random variables with zero mean and finite variance σ_z^2 ; and let

$$J^{(n)}(b) = \sum_{t=1}^n [1(x_{t-1} < b; x_t \geq b) + 1(x_{t-1} > b; x_t \leq b)]$$

denote the normalized number of crossings of level b . If x_0 and z_1 have bounded pdf's with finite variance then we must have:

$$J^{(n)}(b) \xrightarrow{d} \frac{E f_j^2 j g_j}{\sigma_z^2} jZj;$$

where Z is a standard Normal random variable.

Proof. See Theorem 1 in Burrige and Guerre (1996). ■

Lemma 13 (Lévy, 1948) Let Z be a standard Normal random variable and let

$$I_B(0; 1) = \lim_{\epsilon \neq 0} \frac{1}{2\epsilon} \int_0^1 [1[i \pm \epsilon \cdot B(s) \leq \pm \epsilon] ds];$$

where $B(r)$ is a Brownian motion process on $[0, 1]$. Then

$$jZj \stackrel{d}{=} I_B(0; 1);$$

Proof. See Theorem 2.3 in Revuz and Yor (1991). ■

Lemma 14 Let x_t be a stationary Gaussian sequence with covariances c_i satisfying the "Berman condition": $c_i \log i \rightarrow 0$ as $i \rightarrow \infty$. Then all extreme statistics have the same asymptotic distributions as an i.i.d. Gaussian sequence.

Proof. Theorem 2.5.2 in Leadbetter and Rootzén (1988). ■

Lemma 15 If $\{x_t\}_{t=1}^\infty$ is a sequence of i.i.d. random variables then for large n

$$\begin{aligned} E \left[\sum_{t=1}^n x_t^{1/2} J^{(n)} \right] &= O(\log n) \\ \text{Var} \left[\sum_{t=1}^n x_t^{1/2} J^{(n)} \right] &= O(\log n): \end{aligned}$$

Proof. See for instance Port (1994). ■

Lemma 16 Let $\{x_i\}_{i=1}^\infty$ a sequence of random variables such that $\lim_{i \rightarrow \infty} E(x_i) = 1$; and $\lim_{i \rightarrow \infty} \text{Var}(x_i) = 0$. Then

$$x_i \xrightarrow{P} 1:$$

Proof. See for instance Arnold (1990). ■

Lemma 17 Let $x_i = x_{i-1} + z_i$ where $\{z_i\}_{i=1}^\infty$ are continuous i.i.d. random variables with finite variance σ_z^2 and symmetric pdf around a zero mean. If t^0 is the random time of occurrence of the maximum of $\{x_i\}_{i=1}^t$ then for any $u \in [0, 1]$:

$$P(t^0 \leq t \cdot u) = \frac{2}{\pi} \int_0^u \arcsin v \, dv$$

Proof. See Loève (1978). ■

A1. Proof of Proposition 1

Let $x_t = x_{t-1} + z_t$ with z_t satisfying assumptions C1-C4, and let

$$\begin{aligned} a^{(n)} &= n^{1/2} \sum_{t=1}^n 1(R_t^{(x)} > 0) = n^{1/2} J_0^{(n)} \\ &= n^{1/2} \sum_{t=1}^n 1(x_t = x_{t:t}) + n^{1/2} \sum_{t=1}^n 1(x_t = x_{1:t}) \\ &= a_1^{(n)} + a_2^{(n)} \end{aligned}$$

Note that $a^{(n)}$ is the frequency of upper and lower records in the sample x_1, \dots, x_n , and that we could also split this frequency into the sum of the frequencies of upper and lower records as:

$$a_1^{(n)} = \sum_{t=1}^n \frac{n^{i-1} x_t}{n^{\frac{3}{4}}} - \frac{n^{i-1} x_{t;t}}{n^{\frac{3}{4}}} = 0 \quad \frac{t}{n} \rightarrow \frac{t_i}{n}$$

$$a_2^{(n)} = \sum_{t=1}^n \frac{n^{i-1} x_t}{n^{\frac{3}{4}}} - \frac{n^{i-1} x_{1;t}}{n^{\frac{3}{4}}} = 0 \quad \frac{t}{n} \rightarrow \frac{t_i}{n} :$$

Now defining $r = t/n$; where $t = 1; 2; \dots; n$; and letting $n \rightarrow \infty$ we obtain from direct application of lemmas 1, 2 and 3:

$$a_1^{(n)} \rightarrow \int_0^1 \mathbb{1}_{[B(r) > 0]} \sup_{s \in [0;1]} fB(s) g = 0 \quad dr$$

$$a_2^{(n)} \rightarrow \int_0^1 \mathbb{1}_{[B(r) < 0]} \inf_{s \in [0;1]} fB(s) g = 0 \quad dr$$

Finally, it follows from lemma 12 and from the definition of local time that

$$a_i^{(n)} \rightarrow \int_0^1 \mathbb{1}_{[|B(r)| = 0]} dr; \quad i = 1; 2$$

$$= \int_0^1 \mathbb{1}_{[B(r) = 0]} dr$$

$$= 2l_B(0; 1) \lim_{\pm \rightarrow 0} \pm = 0;$$

where $l_B(0; 1)$ represents the local time of the Brownian motion $B(\cdot)$ at zero. Thus the asymptotic probability for the appearance of either upper or lower records must be equal to zero, that is:

$$\lim_{t \rightarrow \infty} P(x_t = x_{t;t}) = 0$$

$$\lim_{t \rightarrow \infty} P(x_t = x_{1;t}) = 0:$$

A2. Proof of Theorem 3

Consider a time series process $x_t = \sum_{i=1}^t \varepsilon_i$ where ε_i are continuous i.i.d. random variables with zero mean and variance σ^2 . Let $y_t = x_t - x_{t;t}$ and $y_t^0 = x_t - x_{1;t}$ and split the RUR test statistic as

$$J^{(n)} = J_1^{(n)} + J_2^{(n)};$$

with

$$\begin{aligned}
J_1^{(n)} &= n^{i-1/2} \sum_{t=1}^n 1(y_t = 0) \\
&= n^{i-1/2} \sum_{t=1}^n 1(y_{t-1} < 0; y_t = 0) + n^{i-1/2} \sum_{t=1}^n 1(y_{t-1} \geq 0; y_t = 0) \\
&= n^{i-1/2} \sum_{t=1}^n 1(y_{t-1} < 0; y_t = 0); \\
J_2^{(n)} &= n^{i-1/2} \sum_{t=1}^n 1(y_t^0 = 0) \\
&= n^{i-1/2} \sum_{t=1}^n 1(y_{t-1}^0 > 0; y_t^0 = 0) + n^{i-1/2} \sum_{t=1}^n 1(y_{t-1}^0 \leq 0; y_t^0 = 0) \\
&= n^{i-1/2} \sum_{t=1}^n 1(y_{t-1}^0 > 0; y_t^0 = 0); \quad \text{since } P(y_{t-1}^0 \leq 0; y_t^0 = 0) = 0;
\end{aligned}$$

where we have used the fact that

$$\begin{aligned}
\sum_{t=1}^n 1(y_{t-1} \geq 0; y_t = 0) &= \sum_{t=1}^n 1(y_{t-1} < 0; y_t = 0) \\
\sum_{t=1}^n 1(y_{t-1}^0 \leq 0; y_t^0 = 0) &= \sum_{t=1}^n 1(y_{t-1}^0 > 0; y_t^0 = 0)
\end{aligned}$$

Notice that the number of lower records of x_t in any given time interval is the same as the number of upper records of y_t in that same interval. Therefore the asymptotic distribution of $J_1^{(n)}$ and $J_2^{(n)}$ must be identical. To obtain this distribution we will proceed by first showing that the time series processes defined as y_t and y_t^0 are asymptotically random walks. By symmetry, the behavior of y_t^0 must be statistically equal to that of y_t . It is therefore enough to study the properties of the process $\{y_t\}_{t=1}^n$:

The conditional variance of y_t given that $x_{t^0:t} = x_{t^0}$ ($t^0 \in [1; n] \setminus Z$) is

$$\text{var}(y_t | x_{t^0:t} = x_{t^0}) = \text{var} \sum_{i=t^0+1}^t \tilde{A}_i^2 = (t - t^0) \frac{1}{2}$$

From lemma 19, the random variable $t^0=t$ has an arcsine distribution with pdf:

$$f(t^0=t) = \frac{2}{\pi \sqrt{1 - (t^0=t)^2}}; \quad t^0=t \in [0; 1];$$

from which we obtain the following expression for the unconditional variance:

$$\text{var}(y_t) = \frac{2\frac{3}{4}^2}{\frac{1}{4}} \int_0^1 \frac{t - t^0}{1 - (t^0=t)^2} d(t^0=t) = t \frac{3}{4}^2:$$

As a consequence, y_t cannot be an $I(0)$ time series process. In fact, if we write $y_t = y_{t-1} + \hat{\epsilon}_t$ where $\hat{\epsilon}_t$ is $I(0)$ and force the equality between this representation and the definition, we get

$$\begin{aligned} \hat{\epsilon}_t &= z_t - \Phi x_{t;t} \\ &= z_t - (x_{t;t} - x_{t-1;t-1}) \\ &= z_t; \text{ if } x_t < x_{t-1;t-1} \\ &= z_t - (x_t - x_{t-1;t-1}); \text{ if } x_t \geq x_{t-1;t-1} \end{aligned}$$

Now, from Proposition 1, we know that the long-run frequency of records is equal to zero, and thus $\lim_{t \rightarrow \infty} P(x_t \geq x_{t-1;t-1}) = 0$: It follows that $\hat{\epsilon}_t = z_t$ with probability $p_t = P(x_t < x_{t-1;t-1}) \rightarrow 1$. In particular:

$$\begin{aligned} E(\hat{\epsilon}_t) &= 0 \\ \text{var}(\hat{\epsilon}_t | x_{t^0} = x_{t;t}) &= \frac{3}{4}^2 \text{ with probability } p_t \rightarrow 1 \\ \text{var}(\hat{\epsilon}_t | x_{t^0} = x_{t;t}) &= (t - t^0 - 1) \frac{3}{4}^2 \text{ with probability } 1 - p_t \rightarrow 0; \end{aligned}$$

from where the unconditional variance of $\hat{\epsilon}_t$ is obtained:

$$\begin{aligned} \text{var}(\hat{\epsilon}_t) &= \frac{2\frac{3}{4}^2}{\frac{1}{4}} \int_0^1 \frac{t - t^0 - 1}{1 - (t^0=t)^2} d(t^0=t) \\ &= \frac{3}{4}^2 - t \frac{1}{2} \text{ with probability } 1 - p_t \rightarrow 0: \\ \text{var}(\hat{\epsilon}_t) &= \frac{2\frac{3}{4}^2}{\frac{1}{4}} \int_0^1 \frac{1}{1 - (t^0=t)^2} d(t^0=t) \\ &= \frac{3}{4}^2 \text{ with probability } p_t \rightarrow 1: \end{aligned}$$

Since in practice it can be assumed that the process x_t was generated at $t = 1$, we conclude that $\hat{\epsilon}_t$ is $I(0)$.

It can also be shown that for t small enough the process y_t has a stochastic unit root. The heuristic reasoning is as follows. Writing $y_t = a_t y_{t-1} + z_t$ and assuming $y_{t-1} \neq 0$ (event whose long-run frequency equals one) we obtain the expected value of the process a_t given the past of y_t :

$$E(a_t | y_{t-1}) = 1 + \frac{\hat{\epsilon}_t - z_t}{y_{t-1}}:$$

Thus there is a possibly non-observable period of time during which $\hat{\epsilon}_t$ can be less than z_t ; implying a transitory short-memory behavior for y_t . Notice

however that as $t \rightarrow 1$ we get $E(a_t | y_{t-1}) \rightarrow 1$; and thus y_t becomes an $I(1)$ process.

Given that y_t is an $I(1)$, and noting that for this process a zero "crossing" amounts to a visit to the origin (crossing over the zero level is impossible), it follows from lemma 14 that

$$J_1^{(n)} \sim \frac{E f_j^2 | j g}{\sqrt{42}} |jZ|;$$

where Z is a standard Normal random variable. From lemma 15 the distribution of $|jZ|$ is the same the local time at zero of a Brownian motion in $[0,1]$, say $I_B(0;1)$. Therefore we can write

$$J_1^{(n)} \sim \frac{E f_j^2 | j g}{\sqrt{42}} I_B(0;1);$$

By the same token we have:

$$J_2^{(n)} \sim \frac{E f_j^2 | j g}{\sqrt{42}} I_B(0;1);$$

Since the pdf of the absolute value of a standard Normal random variable Z is given by

$$f_{|Z|}(u) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{u^2}{2}\right); \quad u \geq 0;$$

we can easily obtain for the pdf of $J_i^{(1)}$ ($i = 1; 2$) the following expression:

$$f_{J_i^{(1)}}(u) = \frac{2}{\sqrt{42} \frac{E f_j^2 | j g}{\sqrt{42}}} \exp\left[-\frac{h^2}{2 \frac{E f_j^2 | j g}{\sqrt{42}}}\right]; \quad h \geq 0; \quad i = 1; 2;$$

Our next step is to show that $J_1^{(n)}$ and $J_2^{(n)}$ are asymptotically independent. This will allow us to invoke lemma 10 and derive the asymptotic pdf of $J^{(n)} = J_1^{(n)} + J_2^{(n)}$ as the auto-convolution of the pdf of the absolute value of a standard Normal random variable:

To show the independence of $J_1^{(1)}$ and $J_2^{(1)}$... first let $T_n = o(n)$ with $\lim_{n \rightarrow \infty} T_n = 1$, and write:

$$J_1^{(n)} = n^{1/2} \sum_{t=1}^{T_n} 1(y_t = 0) + n^{1/2} \sum_{t=T_n+1}^{n} 1(y_t = 0)$$

$$J_2^{(n)} = n^{1/2} \sum_{t=1}^{T_n} 1(y_t^0 = 0) + n^{1/2} \sum_{t=T_n+1}^{n} 1(y_t^0 = 0);$$

It is easy to see that

$$\prod_{t=1}^{T_n} 1(y_t = 0) \xrightarrow{P} 0$$

$$\prod_{t=1}^{T_n} 1(y_t = 0) \xrightarrow{P} 0:$$

That is, for n large enough we get

$$J_1^{(n)} \gg \prod_{t=T_n+1}^{T_n} 1(y_t = 0)$$

$$J_2^{(n)} \gg \prod_{t=T_n+1}^{T_n} 1(y_t^0 = 0):$$

Therefore it is enough to show that the families of binary random variables $f1(y_t = 0)g_t$ and $f1(y_t^0 = 0)g_{t^0}$ are independent for t and t^0 sufficiently large. Let us show that this is the case.

First of all, the random variables $1(y_t = 0)$ and $1(y_t^0 = 0)$ are asymptotically independent for t large enough, since we have:

$$\lim_{t \rightarrow \infty} P(fy_t = 0; y_t^0 = 0) = \lim_{t \rightarrow \infty} P(fx_t = x_{t,t}; x_t = x_{1,t}g)$$

$$= 0 \quad (\text{since the events } fx_t = x_{t,t}g \text{ and } fx_t = x_{1,t}g$$

are mutually exclusive for $t > 1$):

$$= \lim_{t \rightarrow \infty} P(fy_t = 0)P(fy_t^0 = 0) = (0)(0) = 0:$$

$$\lim_{t \rightarrow \infty} P(fy_t = 0; y_t^0 \neq 0) = \lim_{t \rightarrow \infty} P(fy_t = 0) \lim_{t \rightarrow \infty} P(fy_t = 0; y_t^0 = 0) = 0 \cdot 0 = 0$$

$$= \lim_{t \rightarrow \infty} P(fy_t = 0)P(fy_t^0 \neq 0) = (0)(1) = 0:$$

$$\lim_{t \rightarrow \infty} P(fy_t \neq 0; y_t^0 = 0) = \lim_{t \rightarrow \infty} P(fy_t^0 = 0) \lim_{t \rightarrow \infty} P(fy_t = 0; y_t^0 = 0) = 0 \cdot 0 = 0$$

$$= \lim_{t \rightarrow \infty} P(fy_t \neq 0)P(fy_t^0 = 0) = (1)(0) = 0:$$

$$\lim_{t \rightarrow \infty} P(fy_t \neq 0; y_t^0 \neq 0) = \lim_{t \rightarrow \infty} P(fy_t^0 \neq 0) \lim_{t \rightarrow \infty} P(fy_t \neq 0; y_t^0 = 0) = 1 \cdot 0 = 0$$

$$= \lim_{t \rightarrow \infty} P(fy_t \neq 0)P(fy_t^0 \neq 0) = (1)(1) = 1:$$

Let us now show the independence of the variables $1(y_t = 0)$ and $1(y_{t_i}^0 = 0)$ for $i \geq 0$ and t asymptotically large. First of all note that the of events $fy_t = 0g$ and $\bigcap_{j=0}^{i-1} 2_{t_i j} = x_{t,t} \wedge x_{t_i i}$ are equivalent. Thus we can write

$$fy_t = 0g = \bigcap_{j=0}^{i-1} 2_{t_i j} = x_{t,t} \wedge x_{t_i i} :$$

Similarly, we have:

$$y_{t_i}^0 = 0 = f x_{t_i i} = x_{1,t_i} g :$$

Therefore

$$\begin{aligned} \textcircled{c} \quad y_t = 0; y_{t_i}^0 = 0 &= \sum_{j=0}^{\infty} x_{t_i j}^2 = x_{t_i t_i} x_{1; t_i i}, \\ &= \sum_{j=0}^{\infty} x_{t_i j}^2 = \sum_{k=0}^{\infty} \phi x_{t_i k; t_i k} + R_{t_i i}, \end{aligned}$$

But since $fy_t = 0g \frac{1}{2} f\phi x_{t;t} > 0g$; it follows that

$$\textcircled{c} \quad y_t = 0; y_{t_i}^0 = 0 \frac{1}{2} \sum_{j=0}^{\infty} x_{t_i j}^2 > R_{t_i i},$$

and thereby

$$\begin{aligned} \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t = 0; y_{t_i}^0 = 0 &= \lim_{t \rightarrow \infty} \sum_{j=0}^{\infty} x_{t_i j}^2 > R_{t_i i}, \\ &= 0 \quad (\text{by the unboundness of the range sequence}) \\ &= \lim_{t \rightarrow \infty} P \quad y_t = 0g P \quad fy_{t_i}^0 = 0 : \end{aligned}$$

Now we just have to proceed as for $i = 0$: That is,

$$\begin{aligned} \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t = 0; y_{t_i}^0 \neq 0 &= \lim_{t \rightarrow \infty} P \quad fy_t = 0g \quad \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t = 0; y_{t_i}^0 = 0 = 0 \quad i = 0 \\ &= \lim_{t \rightarrow \infty} P \quad y_t = 0g P \quad fy_{t_i}^0 \neq 0 = (0)(1) = 0: \\ \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t \neq 0; y_{t_i}^0 = 0 &= \lim_{t \rightarrow \infty} P \quad y_{t_i}^0 = 0 \quad \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t = 0; y_{t_i}^0 = 0 = 0 \quad i = 0 \\ &= \lim_{t \rightarrow \infty} P \quad y_t \neq 0g P \quad fy_{t_i}^0 = 0 = (1)(0) = 0: \\ \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t \neq 0; y_{t_i}^0 \neq 0 &= \lim_{t \rightarrow \infty} P \quad y_{t_i}^0 \neq 0 \quad \lim_{t \rightarrow \infty} \textcircled{c} P \quad y_t = 0; y_{t_i}^0 \neq 0 = 1 \quad i = 0 \\ &= \lim_{t \rightarrow \infty} P \quad y_t \neq 0g P \quad fy_{t_i}^0 \neq 0 = (1)(1) = 1: \end{aligned}$$

The asymptotic independence of the variables $1(y_{t_i} = 0)$ and $1(y_t = 0)$ for $i > 0$ follows trivially from the above result by the symmetry of the problem. Therefore we conclude that the random variables $J_1^{(1)}$ and $J_2^{(1)}$ are independent. Letting $f_{jzj}(\cdot)$ denote the pdf of the random variable jZj we may finally

invoke lemma 14 to obtain

$$\begin{aligned}
f_{J_0^{(1)}}(v) &= \int_0^v f_{J_1^{(1)}}(w) f_{J_1^{(1)}}(v-i-w) dw \\
&= a^{i-2} \int_0^{\frac{v}{a}} f_{JZJ}\left(\frac{w}{a}\right) f_{JZJ}\left(\frac{v-i-w}{a}\right) dw; \quad \text{where } a = \frac{E f_{JZJ}^2}{\frac{1}{4}} \\
&= a^{i-1} \int_0^{\frac{v}{a}} f_{JZJ}(u) f_{JZJ}\left(\frac{v-i}{a} - u\right) du; \quad \text{where we let } u = \frac{w}{a} \\
&= \int_0^v f_{JZJ}(u) f_{JZJ}(v-i-u) du \\
&= \frac{1}{2} \exp\left(-\frac{v^2+2}{4}\right) [1 - \Phi(v)]:
\end{aligned}$$

To prove the consistency of the test against stationary alternatives satisfying the Berman condition we invoke lemmas 14 and 15, following which $E J^{(n)} \gg \text{Var} f_{J^{(n)}} \gg O(n^{i-2} \log n) \rightarrow 0$ as $n \rightarrow \infty$: Finally, we apply lemma 16 to obtain:

$$J^{(n)} \rightarrow 0$$

A3. Proof of Proposition 6

Letting $x_t = w_t + s_t$; the proof is a straight consequence of the fact that

$$n^{i-2} s_t \rightarrow 0; \quad \text{as } n \rightarrow \infty$$

Now since

$$\begin{aligned}
R_t^{(x)} &= x_{t;t} - x_{1;t} \\
&= w_{t;t} + s_{t;t} - w_{1;t} - s_{1;t};
\end{aligned}$$

we obtain

$$\begin{aligned}
1(R_t^{(x)} > 0) &= 1\left(\frac{1}{4} n^{i-2} R_t^{(x)} > 0\right) \\
&= 1\left(\frac{1}{4} n^{i-2} R_t^{(w)} + \frac{1}{4} n^{i-2} R_t^{(s)} > 0\right):
\end{aligned}$$

Thus

$$\begin{aligned}
J^{(n)} &= n^{i-2} \sum_{t=1}^n 1\left(\frac{1}{4} n^{i-2} R_t^{(w)} > -\frac{1}{4} n^{i-2} R_t^{(s)}\right) \\
&\gg n^{i-2} \sum_{t=1}^n 1\left(\frac{1}{4} n^{i-2} R_t^{(w)} > 0\right) \quad \text{for large enough } n
\end{aligned}$$

since for $0 < t \leq n$ $R_t^{(s)} \cdot R_n^{(s)} = o(n^{i-2})$:

References

- [1] F.M. Aparicio (1995a), "Nonlinear Modelling and Analysis under Long-Range Dependence with an Application to Positive Time Series", PhD thesis. Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne (Switzerland).
- [2] F.M. Aparicio and C.W.J. Granger (1995b), "Nonlinear Cointegration and Some New Tests for Comovements", Discussion Paper 95-15, Department of Economics, University of California at San Diego.
- [3] F.M. Aparicio, A. Escribano and A. Garcia (2000), "Synchronicity between Macroeconomic Time Series: an Exploratory Analysis", Working Paper 00-38, Statistics and Econometric Series, Universidad Carlos III de Madrid.
- [4] F.M. Aparicio, A. Escribano and A. Garcia (2002), "Range Unit Root Tests", Working Paper 03-11, Statistics and Econometric Series, Universidad Carlos III de Madrid.
- [5] S. F. Arnold (1990), Mathematical Statistics, Prentice-Hall, New-Jersey, 1990.
- [6] M. A. Arranz and A. Escribano (1998b), "Detrending Procedures and Cointegration Testing: ECM Tests under Structural Changes", Working Paper, Statistics and Econometric Series, Universidad Carlos III de Madrid.
- [7] H. Anderson and F. Vahid (1998) "Detecting Common Components using Canonical Correlations", Journal of Econometrics, 84, 1-36.
- [8] J. Bai and P. Perron (1998) "Estimating and Testing Linear Models with Multiple Structural Changes", Econometrica, 66, 47-48.
- [9] A. Banerjee, R.L. Lumsdaine and J. Stock (1992), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence", Journal of Business and Economic Statistics, 10, 271-288.
- [10] P. Billingsley (1968), "Convergence of Probability Measures", John Wiley, New York.
- [11] P. Burridge and E. Guerre (1996), "The Limit Distribution of Level Crossings of a Random Walk", Econometric Theory, 12, 705-723.
- [12] O.H. Bustos and V.J. Yohai (1986), "Robust Estimates for ARIMA Models", Journal of the American Statistical Association, 81, 155-168.
- [13] A. Dickey and W.A. Fuller (1979), "Distribution of the Estimators for AR Time Series with a Unit Root", Journal of the American Statistical Association, 74, 427-431.

- [14] P. Embrechts, C. Klüppelberg and T. Mikösch (1999), *Modelling Extremal Events (for Insurance and Finance)*, Springer-Verlag, Heidelberg.
- [15] L. Ermini and C.W.J. Granger (1993), "Generalizations on the Algebra of $I(1)$ Processes", *Journal of Econometrics*, 58, 369-384.
- [16] P.H. Franses and N. Haldrup (1994), "The Effects of Additive Outliers on Tests for Unit Roots and Cointegration", *Journal of Business and Economic Statistics*, 12, 4, 471-478.
- [17] P. Franses and A. Lucas (1997), "Outlier Robust Cointegration Tests", VU Research Memorandum 97-45. Vrije Universiteit.
- [18] W. Feller (1971), "An Introduction to Probability Theory and its Applications" (vol. 2), John Wiley and Sons, New-York.
- [19] W.A. Fuller (1976), *Introduction to Statistical Time Series*, John Wiley and Sons, New-York.
- [20] J. Galambos (1984), "Order Statistics", In *Handbook of Statistics*, eds. P.P. Krishnaiah and P.K. Sen, Amsterdam: North-Holland, Volume 4, pp. 359-382.
- [21] C.W.J. Granger and Halman J.J. (1991), "Nonlinear Transformations of Integrated Time Series", *Journal of Time Series Analysis*, 12, 3, 207-224.
- [22] C.W.J. Granger and T. Terasvirta (1993), *Modeling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- [23] S. Hamori and A. Tokihisa (1997), "Testing for a Unit Root in the Presence of a Variance Shift", *Economic Letters*, 57, 245-253.
- [24] W. Hardle (1991), *Smoothing Techniques: with Implementation in S*, Springer-Verlag, Berlin.
- [25] M.N. Hasan and R.W. Koenker (1996), "Robust Rank Tests of the Unit Root Hypothesis", Working Paper 96-0100, University of Illinois at Urbana-Champaign.
- [26] N. Herrndorf (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", *Annals of Probability*, 12, 141-153.
- [27] H. Hoek, A. Lucas and H.K. Van Dijk (1995), "Classical and Bayesian Aspects of Robust Unit Root Inference", *Journal of Econometrics*, 69, 27-59.
- [28] I. Karatzas and S.E. Shreve (1991), *Brownian Motion and Stochastic Calculus*, Springer, New-York.

- [29] S. Katz, N. Johnson and S.I. Kotz (1985), "Distributions in Statistics", John Wiley & Sons, New-York.
- [30] D. Kwiatkowski, P.D.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159-178.
- [31] M.R. Leadbetter and H. Rootzén (1988), "Extremal Theory for Stochastic Processes", *The Annals of Probability*, 16, 2, 431-478.
- [32] P. Lévy (1948), *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris.
- [33] S. Leybourne, T. Mills and P. Newbold (1998), "Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break under the Null", *Journal of Econometrics*, 87, 191-203.
- [34] G. Lindgren and H. Rootzén (1987), "Extremal Values: Theory and Technical Applications", *Scandinavian Journal of Statistics*, 14, 241-279.
- [35] A.W. Lo (1991), "Long-Term Memory in Stock Market Prices", *Econometrica*, 59, 4, 1279-1313.
- [36] A. Lucas (1995), "An Outlier Robust Unit-Root Test with an Application to the Extended Nelson-Plosser Data", *Journal of Econometrics*, 66, 153-173.
- [37] A. Lucas (1995), "Unit Root Tests Based on M-estimators", *Econometric Theory*, 11, 331-346.
- [38] J.G. Mackinnon (1994), "Approximate Asymptotic Distribution Functions for Unit Root and Cointegration tests", *Econometrica*, 60, 1, 145-157.
- [39] G.S. Maddala and In-Moo Kim (1998), "Unit Roots, Cointegration and Structural Change", Cambridge University Press, Cambridge.
- [40] D. Malinopoulos (2000), "A Note on Nonstationarity, Structural Breaks and the Fisher Effect", *Journal of Banking and Finance*, 24, 695-797.
- [41] R.D. Martin and V.J. Yohai (1986), "Influence Functionals of Time Series", *The Annals of Statistics*, 14, 781-818.
- [42] C.R. Nelson and C.I. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, 10, 139-162.
- [43] D. Peña (2001), "Outliers, Influential Observations and Missing Data" in *A Course in Time Series*, eds. D. Peña, G.C. Tiao and R.S. Tsay, New York: John Wiley.

- [44] P. Perron (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica*, 57, 1361-1401.
- [45] P. Perron (1990), "Testing for a Unit Root in a Time Series with a Changing Mean", *Journal of Business and Economic Statistics*, 8, 70, 153-162.
- [46] P. Perron and T.J. Volgelsang (1992), "Nonstationary Time Series and Level Shifts with an Application to Purchasing Power Parity", *Journal of Business and Economic Statistics*, 10, 3, 301-320.
- [47] P. Perron and G. Rodriguez (2000), "Searching for Additive Outliers in Nonstationary Time Series". Working Paper 0005E, Department of Economics, University of Ottawa.
- [48] P.C.B. Phillips (1987), "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.
- [49] P.C.B. Phillips and P. Perron (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, 75, 335-346.
- [50] S. Port (1994), "Mathematical Probability for Applications", John Wiley & Sons, New-York.
- [51] P. Rappoport and L. Reichlin (1989), "Segmented Trends and Nonstationary Time Series", *Economic Journal*, 99, 168-177.
- [52] D. Revuz and M. Yor (1991), *Continuous Martingales and Brownian Motion*, Springer-Verlag, Berlin.
- [53] T.J. Rothenberg and J.H. Stock (1997), "Inference in a Nearly Integrated Autorregressive Model with Non-Normal Innovations", *Journal of Econometrics*, 80, 269-286.
- [54] I. Sanchez (2003), "Efficient Tests for Unit Roots with Prediction Errors", *Journal of Statistical Planning and Inference*, 113, 341-370.
- [55] P. Schotman and H.K. Van Dijk (1991), "A Bayesian Analysis of the Unit Root in Real Exchange Rates", *Journal of Econometrics*, 49, 195-238.
- [56] B.W. Silverman (1986), "Density Estimation for Statistics and Data Analysis", Chapman and Hall.
- [57] C.A. Sims (1988), "Bayesian Skepticism on Unit Root Econometrics", *Journal of Economics Dynamics and Control*, 12, 436-474.
- [58] J.H. Stock (1994), "Deciding Between $I(1)$ and $I(0)$ Time Series", *Journal of Econometrics*, 63, 105-131.
- [59] J.H. Stock (1990), "A Class of Tests for Integration and Cointegration", Mimeo, Kennedy School of Government, Harvard University.

- [60] T.J. Vogelsang (1990), "Testing for a Unit Root in a Time Series with a Changing Mean", *Journal of Business and Economic Statistics*, 8, 70, 153-162.
- [61] E. Zivot and D.W.K. Andrews (1992), "Further Evidence on the 'Great Crash', the Oil Price Shock and the Unit Root Hypothesis", *Journal of Business and Economic Statistics*, 10, 251-270.

Acknowledgments

The authors are grateful to Daniel Peña, Juan Romo, Esther Ruiz and Ismael Sánchez from Universidad Carlos III de Madrid, to Dag Tjostheim and Jostein Paulsen from the University of Bergen, and to Javier Villarroel from Universidad de Salamanca, for their comments and suggestions. Supported by the Spanish CYCIT under Grant PB98-0030.